

PENERAPAN ALGORITMA APRIORI UNTUK MENEMUKAN HUBUNGAN DATA MURID DENGAN NILAI SEKOLAH

Grand

STMIK Nusa Mandiri
Jl. Kramat Raya No.18 Jakarta Pusat
grandnoowen@gmail.com

ABSTRAK

Murid merupakan salah satu komponen terpenting dari sebuah sekolah, karena tanpa adanya murid maka sekolah tidak dapat melaksanakan kegiatan pembelajaran. Setiap murid pada sebuah sekolah memiliki data yang didapatkan pada saat awal masuk dan juga nilai sekolah yang merupakan hasil pembelajaran. Namun, data-data tersebut belum diketahui keterhubungannya sehingga dibutuhkan sebuah metode untuk mengolahnya. Melalui penelitian ini, sebuah metode *Data Mining* yaitu *Association* dengan Algoritma Apriori digunakan untuk mencari hubungan antara data murid dengan nilai sekolah. Data murid merupakan data awal masuk murid yang terdiri dari tempat dan tahun lahir, agama, dan tahun masuk. Nilai minimal *support* yang digunakan sebesar 20% dan nilai *confidence* merupakan parameter yang menentukan seberapa besar hubungan antara data murid dengan nilai sekolah. Nilai *confidence* terbesar diperoleh dari hubungan antara agama Kristen dengan nilai sekolah $\uparrow 77-82$ sebesar 54,5%. Penelitian ini dilakukan pada SD Kristen Kalam Kudus IV Alam Raya.

Kata kunci: Data Murid, Nilai Sekolah, *Data Mining*, *Association*, Algoritma Apriori, Nilai *support* dan *confidence*.

ABSTRACT

Student are one of the most important component of a school, because without the student, the school can't carry out the learning activities. Each student in a school has data obtained at the start of admission as well as the value of the school as a result of learning. However, these data have not known the connection so that required a method to process it. Through this research, a method of Data Mining is Association with Apriori Algorithm used to find the relation between student data with school value. Student data is preliminary entry data consists of place and year of birth, religion, and year of entry. The minimum value of support used by 20% and the value of confidence is a parameter that determines how big the relationship between student data with school grades. The greatest confidence value was obtained from the relationship between christianity and school grades $\uparrow 77-82$ of 54.5%. This research was conducted at SD Kristen Kalam Kudus IV Alam Raya.

Keywords: Student Data, School Grades, Data Mining, Association, Apriori Algorithm, Support and Confidence Values.

1. PENDAHULUAN

Murid merupakan salah satu komponen terpenting dari sebuah sekolah, karena tanpa adanya murid maka sekolah tidak dapat melaksanakan kegiatan pembelajaran. Menurut KBBI Online, pengertian murid adalah orang / anak yang sedang berguru / belajar / bersekolah. Selain murid, komponen lain yang juga penting dari sekolah yaitu guru sebagai pengajar dan bangunan sekolah yang merupakan tempat berlangsungnya kegiatan pembelajaran.

Setiap murid pada sebuah sekolah memiliki data yang didapatkan pada saat awal masuk dan juga nilai sekolah yang merupakan hasil pembelajaran. Data murid didapatkan dari formulir pendaftaran yang diisi oleh calon murid, sedangkan nilai sekolah

merupakan hasil yang didapatkan dari kegiatan pembelajaran yang berlangsung. Nilai sekolah diberikan oleh masing-masing guru yang mengajar.

Namun, data-data tersebut belum diketahui keterhubungannya sehingga dibutuhkan sebuah metode untuk mengolahnya. Dengan kata lain, data-data tersebut masih merupakan data mentah yang belum siap untuk digunakan. Metode yang dipilih juga harus disesuaikan dengan tipe data yang telah terkumpul, sehingga tujuan utama yaitu untuk menemukan hubungan antara data murid dan nilai sekolah dapat tercapai.

2. METODOLOGI

Dalam sebuah penelitian, dibutuhkan metodologi penelitian untuk menentukan apakah tujuan sebenarnya dilakukannya penelitian(Kothari,

2004). Metodologi penelitian yang digunakan pada studi kasus ini dijelaskan pada bagian-bagian selanjutnya.

Jenis Penelitian

Jenis penelitian menentukan arah dari penelitian akan dilakukan. Hal ini juga berhubungan langsung dengan tujuan yang ingin dicapai pada sebuah penelitian. Jenis penelitian yang digunakan pada studi kasus ini adalah *conclusion-oriented*. *Conclusion-oriented* memberikan kebebasan bagi peneliti untuk menentukan masalah penelitian serta mendesain ulang penyelidikan untuk mengkonseptualisasikan penelitian sesuai keinginan peneliti (Kothari, 2004).

Metode Penelitian

Metode penelitian adalah semua metode yang digunakan oleh peneliti selama proses mempelajari masalah penelitian (Kothari, 2004). Metode penelitian meliputi pengumpulan data, pengolahan data, pencarian solusi, dan evaluasi. penjelasan mengenai metode penelitian yang digunakan akan dijelaskan pada bagian-bagian selanjutnya.

Pengumpulan Data

Data yang dikumpulkan pada penelitian ini, meliputi beberapa hal sebagai berikut:

1. Berdasarkan caranya, maka data yang digunakan pada penelitian ini merupakan data sekunder. Menurut KBBI *Online*, data sekunder merupakan data yang diperoleh seorang peneliti secara tidak langsung dari objeknya, tetapi melalui sumber lain, baik lisan maupun tulis. Dalam penggunaannya, data sekunder harus sesuai dengan penelitian yang dilakukan oleh peneliti (Hox & Boeije, 2005). Data diperoleh dari SD Kristen Kalam Kudus IV Alam Raya. Data berupa softcopy dengan format Ms. Excel (.xls, .xlsx). Data yang dikumpulkan adalah data anak kelas 6 SD.
2. *Dataset* yang digunakan pada penelitian ini merupakan *private dataset* (Aggarwal & Yu, 2007). *Private dataset* adalah dataset yang dapat diambil dari organisasi yang dijadikan sebagai objek penelitian. Objek penelitian ini yaitu SD Kristen Kalam Kudus IV Alam Raya.

Pengolahan Data

Metode yang diusulkan untuk mengolah data-data yang terkumpul yaitu *Association* yang merupakan salah satu metode *Data Mining*. Dan Algoritma Apriori dipilih untuk mengolah data. Alasan dipilihnya metode tersebut karena tujuan pada penelitian ini adalah menemukan hubungan data murid dengan nilai sekolah.

Penelitian menerapkan prinsip pemangkasan Apriori, dimana *itemset* yang tidak sering muncul tidak akan diujikan. Metode berjalan sesuai dengan langkah-langkah sebagai berikut:

1. Pindai *database* sekali untuk mendapatkan 1-*itemset* yang sering muncul.
2. Buat *itemset* kandidat panjang(k+1) dari *itemset* k yang sering muncul.
3. Uji kandidat terhadap *database*.
4. Akhiri proses ketika tidak ada himpunan yang sering muncul atau kandidat yang dapat dibentuk.

Pencarian Solusi

Teknik pencarian solusi yang digunakan adalah *Modelling Existing Solutions*. Teknik ini dipilih karena penelitian ini memodifikasi metode yang sudah pernah digunakan sebelumnya oleh (Pratama et al., 2009) serta mengubah objek penelitian.

Evaluasi

Teknik evaluasi yang digunakan adalah menghubungkan *itemset* data murid dengan nilai sekolah. Nilai yang diujikan yaitu *confidence*.

3. LANDASAN TEORI

Data Murid

Data murid diperoleh dari formulir pendaftaran murid yang diisikan pada saat awal murid mendaftar ke sekolah. Data ini terdiri dari beberapa atribut, yaitu: nama, tempat / tanggal lahir, agama, jenis kelamin, nama orang tua, alamat tempat tinggal, dan sebagainya. Data ini akan diperbarui apabila terjadi perubahan yang dilaporkan oleh murid atau wali murid. Data bersifat rahasia dan hanya dimanfaatkan oleh pihak internal sekolah.

Nilai Sekolah

Nilai sekolah merupakan nilai yang didapat dari hasil perhitungan nilai rapor semester 7-11 (nilai yang diambil dari kelas 4 semester 1 sampai dengan kelas 6 semester 2) dan nilai ujian sekolah. Rumus perhitungan nilai sekolah dijelaskan dibawah ini:

$$NSP = \frac{70\% \times NR}{30\% \times US} \quad (1)$$

NSP = nilai sekolah per pelajaran

NR = nilai rapor dari semester 7-11

US = ujian sekolah

Setelah ditemukan nilai sekolah per pelajaran, maka langkah selanjutnya adalah mencari nilai sekolah secara keseluruhan dengan rumus sebagai berikut:

$$NSK = \frac{NSP_1 + NSP_2 + \dots + NSP_n}{JP} \quad (2)$$

NSK = nilai sekolah keseluruhan
NSP = nilai sekolah per pelajaran
JP = jumlah pelajaran

Nilai sekolah yang telah melalui proses kedualah yang akan digunakan sebagai *itemset*.

Data Mining

Data mining merupakan proses mengekstrak informasi yang menarik dari data yang berjumlah besar sehingga mendapatkan pola dan pengetahuan yang berguna. Informasi yang diekstrak tidak dapat dianggap sepele, implisit, belum diketahui sebelumnya, dan berpotensi berguna (Han, Kamber, & Pei, 2012; Witten, Frank, & Hall, 2011). Proses ini harus otomatis atau lebih sering semi-otomatis (Witten et al., 2011).

Pola yang ditemukan pasti mengandung keuntungan yang dapat diperoleh, seperti: untuk urusan ekonomi. Pola yang ditemukan membantu kita untuk membuat prediksi yang baik bukan sepele. Pola yang ditemukan dapat disebut struktural, karena memiliki bentuk yang dapat diperiksa, dipertimbangkan, dan digunakan untuk informasi di masa depan. Dengan kata lain, membantu menjelaskan data (Witten et al., 2011).

Awal mula hadirnya *Data mining*, karena perlukannya suatu disiplin ilmu untuk mengolah data yang berjumlah sangat besar. Orang dapat menyebut data berjumlah besar ini sebagai kelimpahan data. Kelimpahan data yang tidak diolah menyebabkan munculnya pernyataan "kuburan data". Para pembuat keputusan belum dapat menggunakan data tersebut karena masih merupakan data mentah (Witten et al., 2011).

Data mining merupakan salah satu langkah dari proses *Knowledge Discovery from Data* atau lebih dikenal dengan singkatan KDD. Berikut langkah-langkah dari KDD (Han et al., 2012):

1. *Data cleaning*
Menghilangkan *noise* dan data yang tidak konsisten.
2. *Data integration*
Sumber-sumber data diintegrasikan.
3. *Data selection*
Data yang sesuai dengan tugas analisis diambil dari *database*.
4. *Data transformation*
Data ditransformasikan dan dikonsolidasikan kedalam bentuk yang sesuai dengan melakukan operasi ringkasan dan pengumpulan.
5. *Data mining*
Proses penting dimana metode cerdas diterapkan untuk menemukan pola data.
6. *Pattern evaluation*
Mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan ukuran ketertarikan.
7. *Knowledge presentation*

Visualisasi dan Teknik representasi pengetahuan digunakan untuk menyajikan pengetahuan yang digali kepada pengguna.

Data yang digunakan dalam proses Data mining dapat bersumber dari *Database data*, *Data warehouse*, *Transactional data*, *time-related* atau *sequence data*, *data streams*, *spatial data*, *engineering design data*, *hypertext* dan *multimedia data*, *graph* dan *networked data*, *Web*. Untuk teknologi yang digunakan adalah *statistics*, *machine learning*, *database system* dan *data warehouses*, *information retrieval*. Untuk target penerapannya meliputi: *business intelligence*, *web search engines*. Beberapa masalah utama yang dapat ditemui seperti: *mining methodology*, *user interaction*, *efficiency and scalability*, *diversity of database types*, *data mining and society* (Han et al., 2012).

Terdapat 5 peran utama pada *Data mining*, yaitu: Estimasi, Prediksi, Klasifikasi, Klastering, dan Asosiasi.

Association

Association merupakan aturan yang mempelajari *item* atau atribut yang selalu muncul secara bersamaan (Ye, 2015). Konsep aturan *association* berasal dari *frequent pattern mining* yang mencari pola barang yang sering muncul secara bersamaan pada sebuah himpunan data. *Frequent pattern mining* mengarah pada penemuan asosiasi dan korelasi antar item dalam kumpulan data transaksional atau relasional yang besar. Nilai *support* dan *confidence* merupakan ukuran dari aturan ketertarikan (Han et al., 2012).

Contoh: nilai *support* sebesar 2% berarti bahwa 2 % dari semua transaksi menunjukkan barang A dan B dibeli secara bersamaan. Dan nilai *confidence* sebesar 50% berarti bahwa 50% pembeli barang A juga membeli barang B. Aturan *association* dianggap menarik jika memenuhi ambang batas *minimum support* dan ambang batas *minimum confidence*. Ambang batas ini bisa diatur oleh pengguna atau pakar domain. Secara umum, aturan *association mining* terdiri dari 2 langkah, yaitu (Han et al., 2012):

1. Temukan semua *itemset* yang sering muncul.
Setiap *itemset* memenuhi *minimum support*.
2. Buat sebuah aturan asosiasi yang kuat dari *itemset* yang sering muncul.
Peraturan harus memenuhi dukungan *minimum support* dan *confidence*.

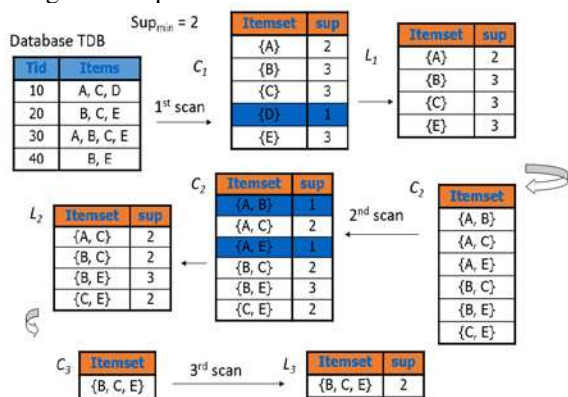
Aturan *association* sering digunakan pada konteks *market basket analysis*. Aturan *association* sering dipakai untuk mengetahui barang yang selanjutnya akan dibeli jika seseorang membeli barang tertentu. Barang yang sering dibeli secara bersamaan akan ditaruh bersama pada toko atau dipasangkan pada *website e-commerce* untuk penjualan barang dan tujuan pemasaran yang lainnya (Ye, 2015). Contoh: seseorang membeli

smartphone pada website e-commerce, maka biasanya orang akan membeli *back cover* juga.

Dari penjelasan diatas, maka dapat disimpulkan bahwa tujuan dari *association* adalah menemukan produk yang berhubungan dengan produk yang ditawarkan maupun dibeli oleh konsumen. Penggunaan *association* juga menyebar luas pada bidang lainnya dengan beberapa perubahan dan perluasan dari aturan awal. Contoh beberapa penerapan aturan *association* lainnya, yaitu analisis teks untuk klasifikasi dan pengambilan dokumen. Selain itu, *association* juga diterapkan pada tipe data yang lain, seperti: data sensus, data linguistik untuk evaluasi penulis, data asuransi, diagnosa medis(Maimon & Rokach, 2010).

Algoritma Apriori

Algoritma Apriori adalah sebuah algoritma untuk menemukan semua kumpulan item (*itemset*) yang memiliki dukungan tidak kurang dari *minimum support*(Wu & Kumar, 2009). Algoritma Apriori diusulkan oleh Agrawal dan Srikant pada tahun 1994 untuk *mining itemset* yang sering muncul untuk aturan asosiasi Boolean(Han et al., 2012). Pada tahap pertama, menghitung semua *itemset* yang ditemukan di *database* transaksi yang dukungannya memiliki *minimum support*. *Itemset* ini disebut *frequent itemset*, meski hanya memiliki beberapa persen minimum ambang batas *support*. Pada fase kedua, minimum ambang batas *confidence* mulai berlaku: Algoritma menghasilkan semua aturan dari *frequent itemset* yang dikonfirmasi melebihi ambang batas(Maimon & Rokach, 2010). Berikut contoh gambaran alur Algoritma Apriori:



Gambar 1. Algoritma Apriori

4. HASIL DAN PEMBAHASAN

Tahap persiapan

Untuk menemukan hubungan data murid dengan nilai sekolah, maka metode yang diusulkan diterapkan pada bagian ini. Data yang diperoleh berupa data murid dan nilai sekolah dari murid kelas 6 SD yang berjumlah 22 orang. Dari data yang telah

terkumpul, hanya dipilih 5 *itemset* yang akan digunakan untuk pengujian.

Tabel 1. Daftar *Itemset*

<i>Itemset</i> Data Murid	<i>Itemset</i> Nilai Sekolah
Data Tempat Lahir	Nilai Sekolah
Data Tahun Lahir	
Data Agama	
Data Tahun Masuk	

Untuk mempermudah perhitungan, maka nilai sekolah dibagi menjadi 4 kelompok, yaitu: nilai 72-77 berinisial A, ↑77-82 berinisial B, ↑82-87 berinisial C, ↑87-92 berinisial D. Tahun lahir menjadi 3 kelompok, yaitu: th 1 = 2000, th 2 = 2001, th 3 = 2002. Dan tahun masuk menjadi 2 kelompok, yaitu: tm 1 = 2007, tm 2 = 2008.

Berdasarkan *itemset* yang telah dipilih seperti tabel diatas, maka data yang terkumpul sebagai berikut:

Tabel 2. Data Murid dan Nilai Sekolah

No	Tempat lahir	Tahun lahir	Agama	Tahun masuk	Nilai sekolah
1	Jakarta	2002	Budha	2008	83,07
2	Jakarta	2000	Kristen	2008	85,32
3	Jakarta	2000	Kristen	2008	72,31
4	Jakarta	2002	Katolik	2008	75,88
5	Dumai	2002	Islam	2008	88,57
6	Jakarta	2001	Kristen	2008	83,96
7	Palembang	2002	Islam	2008	84,45
8	Jakarta	2002	Kristen	2008	79,02
9	Sungailiat	2000	Budha	2007	72,29
10	Jakarta	2000	Katolik	2008	84,47
11	Jakarta	2002	Budha	2008	89,65
12	Jakarta	2002	Kristen	2008	80,38
13	Jakarta	2002	Kristen	2008	90,09
14	Jakarta	2002	Kristen	2008	84,27
15	Jakarta	2002	Budha	2008	79,78
16	Jakarta	2002	Kristen	2007	80,24
17	Jakarta	2001	Kristen	2008	75,75
18	Jakarta	2002	Budha	2008	87,78
19	Jakarta	2002	Budha	2008	82,42
20	Jakarta	2002	Kristen	2008	81,3
21	Jakarta	2002	Kristen	2008	78,31
22	Tangerang	2002	Budha	2008	86,99

Itemset akan dieliminasi apabila memiliki nilai *support* yang kurang dari nilai *minimum support* yang diterapkan. Formula untuk mendapatkan nilai *support per itemset* sebagai berikut(Ye, 2015):

$$support(X) = \frac{| \{S | S \in D \text{ and } S \supseteq X \} |}{N} \quad (3)$$

Pada akhir proses, nilai *confidence* yang akan menentukan seberapa besar hubungan data murid dengan nilai sekolah. berikut formula perhitungan nilai *confidence*(Han et al., 2012).

$$confidence(A \Rightarrow B) = P(B|A) = \frac{support(A \cup B)}{support(A)} \quad (4)$$

Tahap pengujian

1. Data yang diperoleh dipindai terlebih dahulu sehingga menemukan *itemset* yang sering muncul. Pada tahap ini nilai *minimum support* ditentukan sebesar 20%.

Tabel 3. Kandidat *itemset* 1 tempat lahir dengan nilai sekolah.

Itemset	Jumlah	Support
Jakarta	18	81,8%
Dumai	1	4,5%
Palembang	1	4,5%
Sungailiat	1	4,5%
Tangerang	1	4,5%
A	4	18,1%
B	6	27,2%
C	8	36,3%
D	4	18,1%

Tabel 4. Kandidat *itemset* 1 tahun lahir dengan nilai sekolah.

Itemset	Jumlah	Support
th 1	3	13,6%
th 2	2	9%
th 3	17	77,2%
A	4	18,1%
B	6	27,2%
C	8	36,3%
D	4	18,1%

Tabel 5. Kandidat *itemset* 1 agama dengan nilai sekolah.

Itemset	Jumlah	Support
Budha	7	31,8%
Kristen	11	50%
Katolik	2	9%
Islam	2	9%
A	4	18,1%
B	6	27,2%
C	8	36,3%
D	4	18,1%

Tabel 6. Kandidat *itemset* 1 tahun masuk dengan nilai sekolah.

Itemset	Jumlah	Support
tm 1	2	9%
tm 2	20	90,9%
A	4	18,1%
B	6	27,2%
C	8	36,3%
D	4	18,1%

2. Dari proses pindai yang pertama, maka didapatkan kandidat *itemset* 2. Pada tahap ini *itemset* data murid dan nilai sekolah yang masih ada dikombinasikan bersama. Tahap ini juga melakukan hal yang sama sehingga diperoleh sebagai berikut:

Tabel 7. Kandidat *itemset* 2 tempat lahir dengan nilai sekolah.

Itemset	Jumlah	Support
Jakarta, B	6	27,2%
Jakarta, C	3	27,2%
B,C	0	0

Tabel 8. Kandidat *itemset* 2 tahun lahir dengan nilai sekolah.

Itemset	Jumlah	Support
th 3, B	6	27,2%
th 3, C	6	27,2%
B,C	0	0

Tabel 9. Kandidat *itemset* 2 agama dengan nilai sekolah.

Itemset	Jumlah	Support
Budha, Kristen	0	0
Budha, B	1	4,5%
Budha, C	3	13,6%
Kristen, B	6	27,2%
Kristen, C	3	13,6%
B,C	0	0

Tabel 10. Kandidat *itemset* 2 tahun masuk dengan nilai sekolah.

Itemset	Jumlah	Support
tm 2, B	5	22,7%
tm 2, C	8	36,3%
B,C	0	0

3. Pada tahap ini dilakukan pengujian terhadap nilai *confidence* dari kandidat *itemset* yang telah terbentuk. Proses iterasi dihentikan, karena sudah tidak dapat terbentuk lagi gabungan *itemset* data murid dan nilai sekolah.

Tabel 11. Nilai *confidence* tempat lahir dan nilai sekolah.

Itemset	Confidence
Jakarta, B	33,3%
Jakarta, C	33,3%

Tabel 12. Nilai *confidence* tahun lahir dan nilai sekolah.

Itemset	Confidence
th 3, B	35,2%
th 3, C	35,2%

Tabel 13. Nilai *confidence* agama dan nilai sekolah.

Itemset	Confidence
Kristen, B	54,5%

Tabel 14. Nilai *confidence* tahun masuk dan nilai sekolah.

Itemset	Confidence
tm 2, B	25%
tm 2, C	40%

5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, Algoritma Apriori terbukti mampu menemukan hubungan antara data murid dengan nilai sekolah. Nilai *confidence* tertinggi diperoleh dari hubungan itemset agama Kristen dengan nilai sekolah $\uparrow 77-82$ sebesar 54,5%. Proses pindai data hanya terjadi sebanyak 2 kali, dimana pada tahap tersebut telah terbentuk kandidat *itemset* 2. Dari penelitian ini membuktikan bahwa penerapan aturan asosiasi tidak hanya pada *market basket analysis* saja tetapi juga dapat diterapkan pada konteks yang lainnya. Untuk penelitian yang lebih lanjut, maka dapat digunakan itemset data murid yang lebih banyak lagi untuk membentuk kandidat *itemset* yang lebih banyak. Penggunaan algoritma dapat ditambah misalnya dengan Algoritma FP-Growth.

Based on research that has been done, Apriori Algorithm proved able to find relation between student data with school grade. The trust value of Christian itemset relation with school $\uparrow 77-82$ is 54,5%. The data scan process occurs only 2 times, where at this stage have been formed candidate itemset 2. From this research not only on cart analysis criteria can also be applied to other contexts. For further research, more student data items can be used to form more candidate itemset. The use of algorithm can be added with FP-Growth Algorithm.

DAFTAR PUSTAKA

- Arti kata murid. (2017). <https://kbbi.web.id/murid>, diakses 17 September 2017.
- Arti kata data. (2017). <https://kbbi.web.id/data>, diakses 17 September 2017.
- Aggarwal, C. C., & Yu, P. S. (2007). *Privacy-Preserving Data Mining. Security, Privacy and Trust in Modern Data Management*. New York: Springer.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kaufmann.
- Hox, J. J., & Boeijs, H. R. (2005). *Data Collection, Primary vs. Secondary*. *Encyclopedia of Social Measurement*, 5(1), 593-599.
- Kothari, C. R. (2004). *Research Methodology: Methods & Techniques*. New Delhi: New Age International (P) Ltd.
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (Second). Tel-Aviv, Beer-sheva: Springer.
- Pratama, H. C., Bettiza, Matulatan, T. (2009). Data

awal masuk mahasiswa dengan prestasi akademik.

- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining*. Burlington: Morgan Kaufmann series in data management systems.
- Wu, X., & Kumar, V. (2009). *The Top Ten Algorithms in Data Mining*. Boca Raton: CRC Press.
- Ye, N. (2015). *Data Mining Theories, Algorithms, and Examples*. Boca Raton: CRC Press.